

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-108858

(43)Date of publication of application : 12.04.2002

(51)Int.Cl.

G06F 17/28

(21)Application number : 2000-285811

(71)Applicant : INTERNATL BUSINESS MACH
CORP <IBM>

(22)Date of filing : 20.09.2000

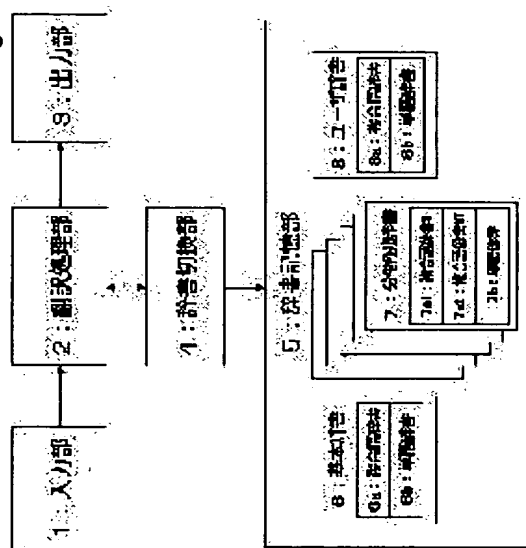
(72)Inventor : HATORI HIROMI
MIYAHIRA TOMOHIRO

(54) MACHINE TRANSLATION METHOD, MACHINE TRANSLATION SYSTEM AND RECORDING MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a translation system capable of being, automatically changed over into more appropriate dictionary.

SOLUTION: The system is provided with a basic dictionary 6 and a field classification dictionary 7 as a dictionary configuration and the field classification dictionary is divided into a compound dictionary T to be a trigger for changing-over the dictionary and a compound dictionary NT not to be the trigger. When a compound word comprised in the compound dictionary T to be the trigger for changing over the dictionary in analyzing an original sentence is detected, the field classification dictionary of the field is made to be in a priority order being higher than the basic dictionary. Besides, the field classification dictionary is subdivided into a main field classification dictionary and a sub-field classification dictionary.



LEGAL STATUS

[Date of request for examination] 24.05.2001

[Date of sending the examiner's decision of rejection] 24.04.2002

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection] 2002-13798

[Date of requesting appeal against examiner's decision of rejection] 23.07.2002

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2002-108858

(P2002-108858A)

(43) 公開日 平成14年4月12日 (2002. 4. 12)

(51) Int.Cl.⁷

G 0 6 F 17/28

識別記号

F I

G 0 6 F 17/28

データベース*(参考)

C 5 B 0 9 1

審査請求 有 請求項の数14 OL (全 14 頁)

(21) 出願番号 特願2000-285811(P2000-285811)

(22) 出願日 平成12年9月20日 (2000. 9. 20)

(71) 出願人 390009531

インターナショナル・ビジネス・マシー
ズ・コーポレーション

INTERNATIONAL BUSIN
ESS MASCHINES CORPO
RATION

アメリカ合衆国10504、ニューヨーク州
アーモンク (番地なし)

(72) 発明者 羽鳥 洋美

神奈川県大和市下鶴間1623番地14 日本ア
イ・ビー・エム株式会社 大和事業所内

(74) 復代理人 100112520

弁理士 林 茂則 (外3名)

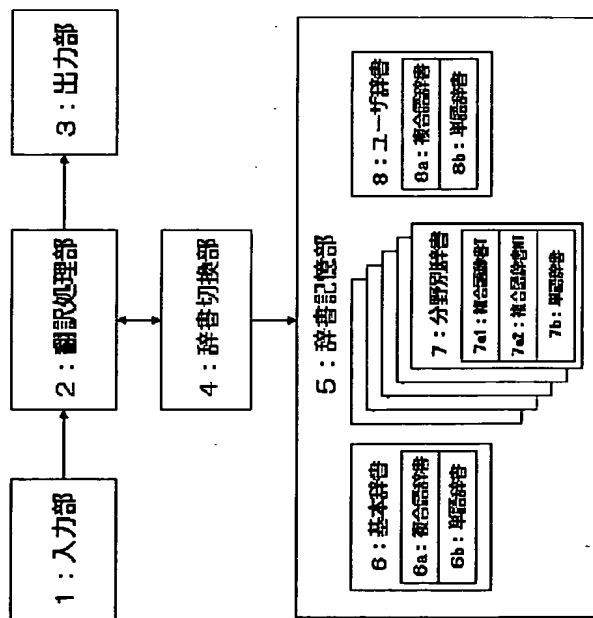
最終頁に続く

(54) 【発明の名称】 機械翻訳方法、機械翻訳装置および記録媒体

(57) 【要約】

【課題】 より適切な辞書に自動切り換えが可能な翻訳システムを提供する。

【解決手段】 辞書構成として基本辞書6および分野別辞書7を有し、分野別辞書を、辞書切り換えのトリガとなる複合語辞書Tとトリガとならない複合語辞書NTに分ける。原文解析において辞書切り換えのトリガとなる複合語辞書Tに含まれる複合語が検出された時にはその分野の分野別辞書を基本辞書より高い優先順位にする。また、分野別辞書をさらに細分し、主分野辞書と副分野辞書に分ける。



【特許請求の範囲】

【請求項 1】基本辞書および少なくとも 1 つの分野別辞書を使用して第 1 言語から第 2 言語への翻訳を行う機械翻訳方法であって、

前記分野別辞書には、辞書の優先度変更のトリガとなる第 1 複合語辞書または第 1 単語辞書と、前記優先度変更のトリガとならない第 2 複合語辞書または第 2 単語辞書と、を含み、

前記第 1 言語の原文を入力するステップと、

入力された前記原文を解析して、前記原文で使用されている複合語または単語を決定するステップと、

決定された前記複合語または単語が前記第 1 複合語辞書または第 1 単語辞書に含まれるかを判断するステップと、

前記判断が真の場合、前記第 1 複合語辞書または第 1 単語辞書が含まれる辞書の優先度を前記基本辞書より高く設定するステップと、

前記優先度が高く設定された状態における辞書を使用して前記第 2 言語の訳文を生成するステップと、を含む機械翻訳方法。

【請求項 2】前記優先度が高く設定される辞書は、前記決定された複合語または単語を含む分野別辞書の第 2 複合語辞書または単語辞書である請求項 1 記載の機械翻訳方法。

【請求項 3】前記分野別辞書には、その分野またはさらに限定された分野に属する主分野辞書および少なくとも 1 つの副分野辞書を含み、

前記主分野辞書および前記副分野辞書には、前記優先度変更のトリガとなる第 1 複合語辞書または第 1 単語辞書と、前記優先度変更のトリガとならない第 2 複合語辞書または第 2 単語辞書と、を含む請求項 1 記載の機械翻訳方法。

【請求項 4】前記優先度が高く設定される辞書は、前記決定された複合語または単語を含む前記主分野辞書または副分野辞書の第 2 複合語辞書または単語辞書、および、前記決定された複合語または単語を含む副分野辞書が含まれる分野に属する主分野辞書の第 2 複合語辞書または単語辞書である請求項 3 記載の機械翻訳方法。

【請求項 5】前記第 1 言語の原文から 1 文を切り出すステップを更に含み、切り出した 1 文毎に前記決定するステップ、前記設定するステップ、および前記訳文を生成するステップを実行し、

前記設定するステップは、前記決定された複合語または単語を含む辞書の優先度を前記基本辞書より高く設定した後、前記優先度が高く設定された辞書に含まれる複合語または単語が所定数の文にわたって使用されなかったときには、前記優先度が高く設定された辞書の優先度を前記基本辞書よりも下げる、請求項 1 記載の機械翻訳方法。

【請求項 6】基本辞書および少なくとも 1 つの分野別辞

書を使用して第 1 言語から第 2 言語への翻訳を行う機械翻訳システムであって、

前記分野別辞書には、辞書の優先度変更のトリガとなる第 1 複合語辞書または第 1 単語辞書と、前記優先度変更のトリガとならない第 2 複合語辞書または第 2 単語辞書と、を含み、

前記第 1 言語の原文を入力する手段と、

入力された前記原文を解析して、前記原文で使用されている複合語または単語を決定する手段と、

決定された前記複合語または単語が前記第 1 複合語辞書または第 1 単語辞書に含まれるかを判断する手段と、前記判断が真の場合、前記第 1 複合語辞書または第 1 単語辞書が含まれる辞書

の優先度を前記基本辞書より高く設定する手段と、

前記優先度が高く設定された状態における辞書を使用して前記第 2 言語の訳文を生成する手段と、を含む機械翻訳装置。

【請求項 7】前記優先度が高く設定される辞書は、前記決定された複合語または単語を含む分野別辞書の第 2 複合語辞書または単語辞書である請求項 6 記載の機械翻訳装置。

【請求項 8】前記分野別辞書には、その分野またはさらに限定された分野に属する主分野辞書および少なくとも 1 つの副分野辞書を含み、

前記主分野辞書および前記副分野辞書には、前記優先度変更のトリガとなる第 1 複合語辞書または第 1 単語辞書と、前記優先度変更のトリガとならない第 2 複合語辞書または第 2 単語辞書と、を含む請求項 6 記載の機械翻訳装置。

【請求項 9】前記優先度が高く設定される辞書は、前記決定された複合語または単語を含む前記主分野辞書または副分野辞書の第 2 複合語辞書または単語辞書、および、前記決定された複合語または単語を含む副分野辞書が含まれる分野に属する主分野辞書の第 2 複合語辞書または単語辞書である請求項 8 記載の機械翻訳装置。

【請求項 10】前記第 1 言語の原文から 1 文を切り出す手段を更に含み、切り出した 1 文毎に前記決定する手段、前記設定する手段、および前記訳文を生成する手段を使用し、

前記設定する手段は、前記決定された複合語または単語を含む辞書の優先度を前記基本辞書より高く設定した後、前記優先度が高く設定された辞書に含まれる複合語または単語が所定数の文にわたって使用されなかったときには、前記優先度が高く設定された辞書の優先度を前記基本辞書よりも下げる、請求項 6 記載の機械翻訳装置。

【請求項 11】辞書の優先度変更のトリガとなる第 1 複合語辞書または第 1 単語辞書と前記優先度変更のトリガとならない第 2 複合語辞書または第 2 単語辞書とを含む少なくとも 1 つの分野別辞書および基本辞書を使用し

10

20

30

40

50

て第1言語から第2言語への翻訳をコンピュータに実行させる、プログラムが記録されたコンピュータ可読な記録媒体であって、コンピュータに、

前記第1言語の原文を入力する機能と、

入力された前記原文を解析して、前記原文で使用されている複合語または単語を決定する機能と、

決定された前記複合語または単語が前記第1複合語辞書または第1単語辞書に含まれるかを判断する機能と、

前記判断が真の場合、前記第1複合語辞書または第1単語辞書が含まれる辞書の優先度を前記基本辞書より高く設定する機能と、

前記優先度が高く設定された状態における辞書を使用して前記第2言語の訳文を生成する機能と、

を実現させるプログラムが記録された記録媒体。

【請求項12】 コンピュータを用いて第1言語から第2言語への翻訳を行わせる際に用いられる辞書が記録された記録媒体であって、

前記辞書には、基本辞書および少なくとも1つの分野別辞書を含み、

前記分野別辞書には、辞書の優先度変更のトリガとなる第1複合語辞書または第1単語辞書と前記優先度変更のトリガとならない第2複合語辞書または第2単語辞書とを含むコンピュータ可読な記録媒体。

【請求項13】 コンピュータを用いて第1言語から第2言語への翻訳を行わせる際に用いられる辞書が記録された記録媒体であって、

前記辞書には、基本辞書および少なくとも1つの分野別辞書を含み、

前記分野別辞書には、その分野またはさらに限定された分野に属する主分野辞書および少なくとも1つの副分野辞書を含み、

前記主分野辞書および前記副分野辞書には、辞書の優先度変更のトリガとなる第1複合語辞書または第1単語辞書と、前記優先度変更のトリガとならない第2複合語辞書または第2単語辞書と、を含むコンピュータ可読な記録媒体。

【請求項14】 第1および第2コンピュータを含むコンピュータシステムまたはコンピュータネットワークにおいて、基本辞書および少なくとも1つの分野別辞書を使用して翻訳役務を提供する機械翻訳方法であって、

前記分野別辞書には、辞書の優先度変更のトリガとなる第1複合語辞書または第1単語辞書と、前記優先度変更のトリガとならない第2複合語辞書または第2単語辞書と、を含み、

前記第1コンピュータが前記第1言語の原文を受信するステップと、

入力された前記原文を解析して、前記原文で使用されている複合語または単語を決定するステップと、

決定された前記複合語または単語が前記第1複合語辞書または第1単語辞書に含まれるかを判断するステップ

と、

前記判断が真の場合、前記第1複合語辞書または第1単語辞書が含まれる辞書の優先度を前記基本辞書より高く設定するステップと、

前記優先度が高く設定された状態における辞書を使用して前記第2言語の訳文を生成するステップと、

前記第2言語の訳文を前記第2コンピュータに送信するステップと、

を含む機械翻訳方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は機械翻訳システムに関する。特に各種辞書の自動切り換えが可能な機械翻訳方法および装置、並びにそのような機械翻訳方法を実行するためのプログラムを記憶したプログラム記憶媒体に関する。

【0002】

【従来の技術】一般に、機械翻訳システムは、一般的な語が登録されている基本辞書およびユーザがユーザ固有の語を登録したユーザ辞書の他に、政治、スポーツ、芸術などの各種の専門分野ごとにその専門用語を登録した1以上の分野別辞書を備えており、それらの辞書を選択的に使用することによって、翻訳処理を実行している。例えば、本出願人が日本で販売している翻訳ソフト「インターネット翻訳の王様」（IBM社の商標）は、基本辞書の他に、「インターネット」、「アート」、「ビジネス」、「スポーツ」、「政治」、および「エンターテインメント」の6つの分野別辞書を備えている。翻訳の質を上げるためには、翻訳に使用する辞書、特に分野別辞書を適切に選択する必要があるが、従来はユーザが原文に合わせて自分で辞書を選択あるいは切り換えるのが普通であった。

【0003】辞書を選択あるいは切り換えを自動的に行う技術も幾つか知られている。辞書の自動切り換えに当たっては、原文の大意を掴んだ上で、適切な分野の辞書を選択することが理想的であるが、大意の把握は容易ではなく、また、話題が転換していく文章においてどの部分を辞書を切り換えるべき単位とするかの決定も難しい。そのため、辞書の自動切り換えを行う現在の一般的な方法は、各辞書を選択するためのキーワードを予め選んでおき、本文中に特定のキーワードが現れたら、そのキーワードに対応する分野別辞書を選択するというものである。

【0004】また、本発明者らによって開発された辞書の自動切り換えの手法に、「パターンベース翻訳システムPalmfreeの訳語選択」、情報処理学会第59回全国大会予稿集、1999、p2-365～p2-366に記載の手法がある。この手法では、辞書をシステム基本辞書（基本辞書）と分野別辞書で構成し、さらに基本辞書と分野別辞書を各々複合語と単語に分け、単語の優先度

を上げるトリガに複合語を用いる。このような手法を用いることにより、より適切な訳語選択が可能になっている。

【0005】なお、特開平8-166955号公報には、原語と訳語の対のリストを与えて、それにどの程度マッチするかによって複数の分野別辞書の優先順位を決定する手法が開示されている。また、特開平7-141375号公報には、複数の分野別辞書を使用し、あらかじめユーザが静的に優先度を指定する手法が開示されている。さらに、特開平5-61902号公報には、キーワードによって原文の分野を特定して、システム辞書中の訳語の順位を変更する手法が開示されている。

【0006】

【発明が解決しようとする課題】分野別辞書を用いて訳語を生成した場合、その後この分野別辞書の優先度を上げることにより、よりの確かな訳文を生成することは可能である。しかし、この手法を用いる場合であっても以下のような改善点がある。

【0007】基本辞書には、一般に使用頻度の高い単語、複合語が多く登録される。基本辞書に登録される一般的な複合語には、たとえばスポーツチームの名称や選手名、映画タイトル等の固有名詞があげられる。これら固有名詞は、たとえばスポーツ分野やエンターテインメント分野に固有の語であり、辞書の自動切り換えの有効なトリガになり得るものである。ところが、これら固有名詞等分野固有の語が一般的に良く知られている（使用頻度が高い）がゆえに基本辞書に登録され、適切な分野別辞書に登録されていないことが多い。このため効果的な辞書の切り換えが働かず、翻訳処理中に期待する辞書に切り換わらない不具合がある。

【0008】また、分野別辞書には、その分野固有の表現で登録される。たとえば英文で「this season」と表現される場合、スポーツ分野では「今シーズン」と日本語訳するのが適切である。よってスポーツ分野の複合語辞書には「this season = 今シーズン」と登録される。ところが、本来「this season」は一般的な語であり、基本辞書においては「this season = この季節」と登録されるべきである。ここで仮にスポーツ分野の話題ではなく、一般的な文書を翻訳している時に「this season」の複合語が入力されれば、これがトリガとなってスポーツ分野の優先度があがり、本来基本辞書を用いるべきところでスポーツ分野の分野別辞書が用いられることになる。これは翻訳中に予期せぬ辞書に切り換わることを意味し、訳文が適切でなくなる不具合を生じる。このような一般的な語であって、かつ特定の分野に固有の表現も存在することは特に日英間の問題に限らず、言語間の語彙が概念的には完全に一致しない翻訳における本質的な問題である。たとえば仏語において「chateau」と表現される場合、一般的には英語では「castle」であるが、ボルドー地方について表現されている時には「winery」

を意味する。このようにある言語において同一の表現であっても、他の言語では異なる別の適切な表現が存在することは常である。よって、このような同一表現複数語義に対する訳語は辞書切り換えのトリガには用いるべきではない。なお、仏語における「chateau」の表記は、本来「ch」のあとの「a」に「^」（アクセント記号）が付けられる。ここでは、文字コードの関係で単に「a」と表記する。以下本明細書において同様である。

10 【0009】さらに、従来技術の分野構成が適切な訳語の選択範囲に合致しないという不具合がある。たとえばスポーツ分野において「ショット」あるいは「シュート」に対応する英単語に「shot」がある。日本語においては慣習的に「ショット」はたとえばゴルフ用語として、「シュート」はサッカーあるいはバスケットボール用語として用いられている。より適切な訳文を生成するには、これら訳文を生成できる辞書を用意する必要がある。

20 【0010】本発明の目的は、より適切な辞書に自動切り換えが可能な翻訳システムを提供することにある。本発明の他の目的は、より詳細な訳出に対応できる辞書構成を有する翻訳システムを提供することにある。さらに、本発明は、より訳質の高い翻訳文を生成できる翻訳システムを提供することにある。

【0011】

30 【課題を解決するための手段】本発明の概要を説明すれば以下の通りである。すなわち、本発明の機械翻訳システムにおいては、その辞書構成として基本辞書および分野別辞書を有し、分野別辞書を、辞書切り換えのトリガとなるデータ（辞書）とトリガとならないデータに分ける。そして、トリガとなるデータには、真にその分野に固有の語を登録する。一方トリガとならないデータには、一般的な用語でもあり、特定分野の用語でもある語を各々の辞書に適した訳語をあてて登録する。このようにトリガになるデータとならないデータを分けて登録することにより、意図しない辞書の切り換えを防止できる。

40 【0012】また、本発明の翻訳システムは、分野別辞書をさらに細分化し、主分野辞書と副分野辞書に分ける。分野別辞書を細分化することにより、さらにきめ細やかな訳出に対応して訳質を向上できる。なお、主分野辞書および副分野辞書を前記トリガになるデータとならないデータを分けて登録できることは勿論である。

50 【0013】なお、トリガになるデータとならないデータに分ける対象は複合語辞書であることが好ましい。単語は同一の綴りが複数分野に存在することが多いので分野判定には不向きであり、複数単語からなる複合語は特定分野で用いられることが多いので分野判定を誤る確率が低くなるためである。しかし単語辞書をトリガになるデータとならないデータに分けることを排除するもので

はない。

【0014】また優先度が上げられる対象は単語辞書またはトリガにならないデータ（複合語辞書）とし、トリガになるデータの優先度はそのまま維持することができる。同じ複合語が特定の分野別辞書および基本辞書の両方に登録されていた場合には、基本辞書の複合語を使用した方が好ましいと思われるからである。もしそのような考慮が不要であれば、単語と複合語両方の優先度を基本辞書より高くしても差し支えない。

【0015】なお、本明細書において、「複合語」とは、2以上の単語からなる有意語のことを意味する。

【0016】

【発明の実施の形態】以下、本発明の実施の形態を図面に基づいて詳細に説明する。ただし、本発明は多くの異なる態様で実施することが可能であり、本実施の形態の記載内容に限定して解釈すべきではない。なお、実施の形態の全体を通して同じ要素には同じ番号を付するものとする。

【0017】以下の実施の形態では、主に方法またはシステムについて説明するが、当業者であれば明らかなとおり、本発明は方法、システムその他、コンピュータで使用可能なプログラムコードが記録された媒体としても実施できる。したがって、本発明は、ハードウェアとしての実施形態、ソフトウェアとしての実施形態またはソフトウェアとハードウェアとの組合せの実施形態をとることができる。プログラムコードが記録された媒体としては、ハードディスク、CD-ROM、光記憶装置または磁気記憶装置を含む任意のコンピュータ可読媒体を例示できる。

【0018】本実施の形態で利用できるコンピュータシステムには、中央演算処理装置（CPU）、主記憶装置（メインメモリ：RAM（Random Access Memory））、不揮発性記憶装置（ROM（Read Only Memory））等を有し、これらがバスで相互に接続される。バスには、その他コプロセッサ、画像アクセラレータ、キャッシュメモリ、入出力制御装置（I/O）等が接続される。バスには、適当なインターフェイスを介して外部記憶装置、データ入力デバイス、表示デバイス、通信制御装置等が接続されてもよい。その他、一般的にコンピュータシステムに備えられるハードウェア資源を備えることが可能なことは言うまでもない。外部記憶装置は代表的にはハードディスク装置が例示できるが、これに限られず、光磁気記憶装置、光記憶装置、フラッシュメモリ等半導体記憶装置も含まれる。なお、データの読み出しのみに利用できるCD-ROM等の読み出し専用記憶装置もデータあるいはプログラムの読み出しにのみ適用する場合には外部記憶装置に含まれる。データ入力デバイスには、キーボード等の入力装置、マウス等ポインティングデバイスを備えることができる。データ入力デバイスには音声入力装置も含む。表示装置としては、CRT、液晶表示

装置、プラズマ表示装置等が例示できる。本実施の形態のコンピュータシステムには、パーソナルコンピュータ、ワークステーション、メインフレームコンピュータ等各種のコンピュータが含まれる。

【0019】本実施の形態の処理システムは、単一のコンピュータシステムとして適用することが可能であるが、複数のコンピュータシステムのネットワークとして適用することも可能である。この場合、コンピュータシステム間の通信にインターネット、LAN、WAN等を用いることができる。これら接続に用いられる通信回線は、専用線、公衆回線の何れでも良い。

【0020】複数のコンピュータシステムで本発明を実現する場合、各コンピュータシステムで利用されるプログラムは、他のコンピュータシステム（例えばサーバ・コンピュータ）に記録されていても良い。つまり、コンピュータシステムで利用する一部のプログラムをリモートコンピュータ（例えば他のサーバコンピュータ又はクライアント・コンピュータ）で分散的に処理または実行できる。なお、他のコンピュータシステムに記録されたプログラムをアドレスで参照する場合には、DNS、URL、IPアドレス等を用いることができる。また、本実施の形態で説明される辞書データは分散的に記録されていても良い。本明細書で辞書内に他の辞書を「含む」と表現している場合には、データ構造上これを含む必要はなく、あるデータ（辞書）と他のデータ（辞書）が概念的に関連付けられている場合を含む。分散的に記録されているデータ（辞書）は前記同様DNS、URL、IPアドレス等を用いて存在場所を特定しアクセスできる。

【0021】なお、インターネットには、イントラネットおよびエクストラネットも含むものとする。インターネットへのアクセスという場合、イントラネットやエクストラネットへのアクセスをも意味する。コンピュータネットワークという用語には、公的にアクセス可能なコンピュータネットワークと私的なアクセスしか許可されないコンピュータネットワークとの両方が含まれるものとする。

【0022】（実施の形態1）本発明に従う機械翻訳システムの構成を図1に示す。以下で説明する実施形態では、機械翻訳システムは英語から日本語への翻訳を実行するが、本発明はそれに限らず、任意の2カ国語間の翻訳に適用可能なものである。本実施の形態のシステムは、翻訳すべき第1言語（英語）の原文を入力するための入力部1、入力された原文から第2言語（日本語）の訳文を生成する翻訳処理部2、この翻訳処理部2で使用する辞書を自動的に切り換えるための辞書切換部4、翻訳処理部2で使用する複数の辞書を記憶している辞書記憶部5、および翻訳処理部2で生成された訳文を出力する出力部3を具備している。また、辞書記憶部5には、基本辞書6、分野別辞書7、ユーザ辞書8を含む。

【0023】入力部1は、原文のテキストを翻訳処理部2へ入力できるものであれば、キーボード、文字認識装置、音声認識装置、インターネットのウェブ・ページ画面、など任意のものでよい。翻訳処理部2は、従来からある通常の機械翻訳エンジンでよい。例えば、そのような翻訳エンジンの一例が、K. Takeda "Pattern-Based Context-Free Grammar for Machine Translation", Proc. of 34th ACL, pp.144-151, 1996およびK. Takeda "Pattern-Based Machine Translation", Proc. of 16th Coling, Vol.2, pp.1155-1158, 1996に記載されている。

【0024】辞書切換部4は本発明の骨子を成すものであり、後述のように、辞書記憶部5（例えばハード・ディスク・ドライブ）に記憶されている分野別辞書の優先度を自動的に切り換える。辞書記憶部5は、基本辞書6、分野別辞書7、およびユーザ辞書8を記憶している。

【0025】基本辞書6は、一般的な語が登録される辞書であり、特定の分野に固有の語を登録しないようにする。複数の分野に渡って一般的に用いられる語は基本辞書6に登録する。基本辞書6は複合語辞書6aおよび単語辞書6bを含む。複合語辞書は、前記の通り複数の単語からなり、それ自体有意な語である。複合語には変数が含まれる。たとえば「to carry out ~」の「~」は変数であり、日本語には「~を実行する」が当てられる。

【0026】分野別辞書7は特定の専門分野の語が登録され、一般に複数存在する。但し単一の分野別辞書で構成されても良い。分野別辞書7には、複合語辞書7a1と複合語辞書7a2と単語辞書7bとを含む。複合語辞書7a1は辞書切り換えのトリガとなる辞書であり、複合語辞書7a2は辞書切り換えのトリガとならない辞書である。複合語辞書7a1と複合語辞書7a2の機能は後に説明する。

【0027】ユーザ辞書8は、ユーザ固有の語が登録されているユーザによって定義され生成される辞書である。本発明においてはユーザ辞書8は任意である。

【0028】出力部3は、翻訳処理部2で生成された訳文を所与の形式、例えば音声、ディスプレイ、印刷、ファイル記憶、などの形式で出力する。

【0029】図1の機械翻訳システムにおける翻訳処理の流れを図2に示す。開始ステップ10で処理が開始され、ステップ11で、入力部1から英語の原文が入力される。次に、ステップ12で、入力された原文から1文が切り出される。英語の場合、(1)単語の最後がピリオドで、次の語の先頭が大文字のとき、または(2)単語の最後が感嘆符、コロン、もしくはセミコロンの場合に、文の区切りとする。ただし、条件(1)が満たされても、例えば「Mr.」のように、文末には現れない表現もあるので、そのような表現をデータとして持ち、原文中の語をそれらの表現と比較して、一致しない場合に文

の区切りとする。また、ピリオドを挟んでその両側に数字があった場合、ピリオドの直後にスペースがあれば、そこで文を区切り、スペースがなければ、ピリオドを小数点とみなして、文を続ける。

【0030】1文が切り出されると、次のステップ13でその1文の原文解析が行われる。原文解析13では、切り出された1文のどの部分が主語で、どの部分が動詞であるかなどの解析が行われ、文の構造が決定される。その際に文法ルールと共に、すべての辞書中の複合語も文の各要素を決定するために使用される。原文の解析が終わると、ステップ14に進んで、分野別辞書の優先度を変更される。優先度の変更については後に詳述する。次にステップ15で、単語の優先度を変更された分野別辞書を使用して訳文が生成される。優先度の高い分野別辞書の単語の訳語を使用することにより、適切な訳語が得られる。次のステップ16で、訳文が生成された文がステップ11で入力された原文における最後の文かどうかチェックされ、最後の文でなければステップ12に戻り、最後の文であれば終了する(ステップ17)。

【0031】次に、実際の原文を例に挙げて、図2の流れを説明する。以下の例は、インターネットのウェブ・ページの翻訳を想定しており、分野別辞書7として次の6つの辞書が使用されるものとするが、もちろん本発明はウェブ・ページ以外の翻訳および以下とは異なった分野別辞書の使用も可能である。

【0032】分野別辞書7

- ・インターネット一般
- ・アート
- ・ビジネス
- ・スポーツ
- ・政治
- ・エンターテイメント

【0033】上記の6種類の分野別辞書の他に、辞書記憶部5には、基本辞書6およびオプションとしてのユーザ辞書8が記憶されており、その優先度は高いものから順に下記のように初期設定されているものとする。

【0034】1. ユーザ辞書

2. 基本辞書
3. インターネット一般
4. アート
5. ビジネス
6. スポーツ
7. 政治
8. エンターテイメント

【0035】3から8までの分野別辞書の優先順位は便宜上のものであり、任意に設定することができる。例えば、翻訳エンジンを初期化したときに見つかった辞書の順番でもよい。初期状態では、すべての分野別辞書の優先度が基本辞書より下になっているが、ユーザにより選択された特定の分野別辞書(例えばスポーツ分野辞書)

の優先度を基本辞書より上に設定することは可能である。その場合、選択されなかった分野別辞書の優先度は基本辞書より下のままである。優先度の管理方法は任意でよい。例えば、辞書名等の辞書識別情報を優先度の順にリストに並べておいてもよい。そのような優先順位リストは、図2のフローを実行する計算機システムの所定のメモリ領域（例えばメインメモリの一部）に保持しておくことができる。

【0036】上記のような辞書を用いた機械翻訳に際し、ステップ11で次の2文が入力されたとする。ここで、"Hiro Hatori"及び"Shige Hayashi"は著名なゴルファーであるとする。

【0037】"Hiro Hatori led the Tour at 30.5%. On par-5 holes Shige Hayashi made a birdie an amazing 49.5% of the time."

【0038】次の1文入力ステップ12では、前述の条件(1)または(2)を満たすところで文を区切ることにより、1文を切り出す。上記の例以下の1文が切り出される。

【0039】"Hiro Hatori led the Tour at 30.5%." 20

【0040】次に、ステップ12で切り出した1文に対し、上述の1から8までのすべての辞書を用いて、原文解析ステップ13が実行される。よく知られているように、各辞書中には単語および複合語があり、更にそれぞれに対応する品詞情報が格納されている。また、

- ・名詞は1つで名詞句を構成することができる
- ・動詞は1つで動詞句を構成することができる
- ・動詞句と名詞句で再び動詞句を構成することができる
- ・名詞句と動詞句で文を構成することができる

といった文法ルールも基本辞書中に格納されている。原文解析ステップ13はそれらの情報を用いて、図3に示すような構文木を作成する。このような構文木の作成方法は周知であるから、作成方法の詳細については省略する。

【0041】図3に示されている略語の意味は下記の通りである。

【0042】

DET	冠詞
N	名詞
NP	名詞句
PP	前置詞句
PREP	前置詞
PUNCT	句読点
S	文
V	動詞
VP	動詞句

【0043】図3の構文木は、文Sが名詞句NP、動詞句VPおよび句読点PUNCTから成っていることを示している。文Sから出ている3本の枝のうち左の枝、すなわち最初の名詞句NPは2つの名詞"Hiro"および"Hat 50

ori"から成っている。この名詞句は複合語であり、本例では、スポーツ分野辞書にのみ登録されている。構文木の真ん中の枝、すなわち名詞句NPに続く動詞句VPは、第2の動詞句VPおよび前置詞句PPから成っている。第2の動詞句VPは、1つの動詞"led"により構成される動詞句VPと、冠詞"the"および名詞"Tour"から成る名詞句NPとから成っている。前置詞句PPは、前置詞"at"と、2つの名詞"30.5"および"%"から成る名詞句NPとから成っている。右の枝にある最後の句読点PUNCTはピリオド"."である。

【0044】原文の解析が終わると、辞書優先度変更ステップ14に進む。上記の例では、"Hiro Hatori"という複合語がスポーツ分野辞書から使用されたので、後に詳しく説明するように辞書の優先順位が変更される。

【0045】次の訳文生成ステップ15では、この変更された優先順位リストに従って訳文の生成が行われる。最初の複合語"Hiro Hatori"は、他の辞書に同じ複合語がないので、スポーツ分野辞書の複合語「ヒロ・ハトリ」という訳語が使用される。次に、"Tour"という単語には、スポーツ分野辞書からの「ツアー」という訳語が使用される。ここで基本辞書にある"tour"という単語に対する「旅行」が使用されないのは、スポーツ分野辞書の優先度が基本辞書より高いためである。優先度の変更については後述する。最終的に以下の訳文が生成される。

【0046】「ヒロ・ハトリは30.5%のツアーを導きました。」

【0047】上記の文は最後の文ではないので、ステップ16からステップ12に戻り、次の文"On par-5 holes Shige Hayashi made a birdie an amazing 49.5% of the tour."が切り出されて、同様な処理が繰り返される。この繰り返しにおいては、スポーツ分野辞書に固有の複合語は検出されないが、後で図4を参照して説明するように、スポーツ分野辞書の単語の優先度はまだ基本辞書よりも高い状態に保たれている。従って、例えば、スポーツ分野辞書および基本辞書の両方に登録されている単語"birdie"に対しては、スポーツ分野辞書からの訳語「バーディ」が使用され、基本辞書にある訳語「小鳥」は使用されない。他の単語についても同様である。2回目の繰り返しで生成される訳文は次の通りである。

【0048】「パー5のホールで、シゲ・ハヤシは、当時の驚くような49.5パーセント(で)バーディを出しました。」

【0049】上記の例では、原文は2文だけであるから、2回目の繰り返しで最後の文に達し、従って2番目の訳文が生成された後、終了する(ステップ17)。なお、これらの訳文は、いずれも本発明に従う機械翻訳システムが実際に出力したものである。2番目の訳文における「(で)」は、ユーザが必要に応じて修正すればよい。

【0050】次に、図4を参照しながら、分野別辞書の優先度変更処理(図2のステップ14に対応)の詳細について説明する。図4のプロセスは各分野別辞書毎に実行される。まず最初のステップ21で、図3に示したような原文解析結果を見て、当該分野別辞書の辞書切り換えのトリガになる複合語辞書Tにある複合語がヒットした(使用された)かどうかをチェックする。ヒットした場合には、ステップ22に進み、その分野別辞書の優先度が既に基本辞書よりも高くなっているかどうかを、例えば前述の優先順位リストを見てチェックする。既に高優先度になっていれば、ステップ24に進み、さもなければステップ23に進む。ステップ23では、当該分野別辞書の優先度を基本辞書よりも上げる。上記の例では、「HiroHatori」という、スポーツ分野辞書にある複合語がヒットしているので、スポーツ分野辞書の優先度が基本辞書より上にされていた。次のステップ24では、未ヒット文数、すなわち当該分野別辞書にある複合語がヒットしなかった文の数が0にリセットされ、その後図2のステップ15に進む。この未ヒット文数は、分野別辞書毎に設定され、その辞書の複合語が最後にヒットしてから何文の間ヒットしていないかをカウントするものである。これは、特定の分野別辞書の複合語が所定数(例えば10)の連続する文の翻訳でヒットしなかった場合に、当該専門分野の話題が終わったものとして、その分野別辞書の優先度を基本辞書より下げるために使用される。

【0051】ステップ21で分野別辞書の複合語辞書T(辞書切り換えのトリガになる複合語辞書)がヒットしなかった場合は、ステップ25に進む。ステップ25はステップ22と同じであり、当該分野別辞書の優先度が既に基本辞書よりも高くなっているかどうかをチェックする。もし高くなっていなければ、何もせずに図2のフローのステップ15に戻る。基本辞書より高優先度になっていると、ステップ26に進んで、上述の未ヒット文数を1だけ増分する。次に、ステップ27で、増分された未ヒット文数が所定の設定値(例えば10)より大きいかどうかをチェックする。未ヒット文数が設定値以下であれば図2のフローのステップ15に戻り、さもなければステップ28に進んで、当該分野別辞書の優先度を基本辞書よりも下げる。

【0052】上記ステップをさらに図5および図6を参照しながら具体的に説明する。なお、各辞書の優先順位は図の左端に示すように上にあるほど高いことを示している。たとえば、図5に示すように、野球分野(スポーツ分野)の複合語辞書NT(辞書切り換えのトリガにならない複合語辞書)に「this season = 今シーズン」というデータがあったとする。そして、「this season」は基本辞書の複合語には登録されず、各々単語として「this = この」「season = 季節」が基本辞書の単語辞書に登録されていたとする。スポーツ分野の辞書の優先

順位が低い状況で、たとえば「People enjoy cherry trees this season.」の一文が入力されるとする。従来技術のように全ての複合語辞書がトリガとなる時には、入力文中の「this season」の複合語によりスポーツ分野の辞書の優先順位が基本辞書より高くなり、「人々は、今シーズン桜を楽しみます。」のように訳出されてしまう。しかし、本実施の形態では、「this season = 今シーズン」が登録されている複合語辞書NTは辞書切り換えのトリガにならないので、たとえ「this season = 今シーズン」がスポーツ分野の辞書に登録されていても、スポーツ分野の辞書の優先順位は上がらず、前記入力文は、「人々は、この季節桜を楽しみます。」のように適切に訳出される。

【0053】一方、入力文が「Tom Miyahira is a home run batter. He hit fifty home runs this season.」の場合(ここでTom Miyahiraは著名な野球選手であるとする)、複合語「Tom Miyahira」がトリガになる複合語辞書Tに登録されていれば、図6に示すように、入力文の第1文中「Tom Miyahira」によってスポーツ辞書の複合語辞書NTと単語辞書が基本辞書より優先度が高くなり、第2文の「this season」は優先度が高くなったスポーツ分野の複合語辞書NTによって「今シーズン」と訳され、「トム・ミヤヒラは、ホームラン打者です。彼は、今シーズン50本のホームランを打ちました。」と適正に訳出される。

【0054】このように、本実施の形態では、辞書切り換えのトリガになるデータとならないデータを分けて記録することにより、前記「this season」のような一般的にもまた特定分野にも用いられる語(英語)を最適な辞書に切り換えて適正な対応する語(日本語)に訳出できる。なお、ここでは英文から日本文への翻訳を説明しているが、その逆にも勿論適用が可能である。たとえば「～を処理する」という日本語に対し、英語の基本辞書には「to deal with～」と登録する一方、コンピュータ分野の分野別辞書には「to process～」、化学分野の分野別辞書には「to treat～」と登録できる。あるいは、「～を実行する」という日本語に対し、基本辞書では「to carry out～/to practice～」、コンピュータ分野では「to run～」と登録できる。あるいは、「～を保存する」という日本語に対し、基本辞書では「to preserve～/to conserve～/to keep～」、コンピュータ分野では「to store～」と登録できる。あるいは、「手を打つ」という日本語に対し、基本辞書では「to take measures」、ビジネス分野では「to strike a bargain」という英語を対応付けることができる。

【0055】さらに、日英間の翻訳の適用に限られないことも勿論である。たとえば「chateau」という仏語に対し、英語の基本辞書では「castle」、ボルドー地方に関する分野別辞書では「winery」を対応付けることができる。

【0056】（実施の形態2）本実施の形態のシステムは、実施の形態1のシステムとほぼ同様である。但し、実施の形態1における分野別辞書7の各々が、図7に示すように、主分野辞書36と副分野辞書37で構成される。副分野辞書37は一般的に複数有するが、1つでも良い。

【0057】主分野辞書36および副分野辞書37の各々には、実施の形態1の分野別辞書と同様に、辞書切り換えのトリガになる複合語辞書Tと、辞書切り換えのトリガにならない複合語辞書NTと、単語辞書を有する。複合語辞書Tおよび複合語辞書NTの機能は実施の形態1と同様である。

【0058】本実施の形態の辞書構成の一例を図8に示す。基本辞書については実施の形態1と同様である。たとえばスポーツ分野においては、「野球」「ゴルフ」等のさらに細分化された分野を有する。このように細分化された分野に副分野辞書を対応付ける。たとえば野球の副分野辞書には、野球固有の語を登録する。「Tom Miyahira」、「home run」等は野球特有の用語であり、また辞書の自動切り換えに適した語である。よって、「Tom Miyahira」「home run」等はスポーツ分野の野球（副分野）の複合語辞書T（辞書の自動切り換えのトリガになる）に登録する。一方、野球固有の表現であるが一般的にも用いられる用語、たとえば「base line」は野球用語においては「ベースライン」であり、一般的には「基準線」であるから辞書の自動切り換えのトリガにするべきではない。よって、これを複合語として登録する場合には野球の副分野辞書の複合語辞書NTに登録する。ところで、スポーツ分野で用いられる語であるが特定の副分野に分類できない語も存在する。たとえば「this season」がそうである。「this season = 今シーズン」は野球に限らず、あるいはゴルフその他の特定のスポーツに限らず、スポーツ全般に用いられる語である。このような語についてはスポーツ全般としてスポーツ分野の主分野辞書に分類する。

【0059】このように分野が階層的に構成された分野別辞書を用いた処理方法は、実施の形態1とほぼ同様である。但し、図9に示すように、原文解析で抽出された複合語がヒットした時に優先順位が上位に変更されるのは分野別辞書の全体ではなく、そのヒットした複合語を含む複合語辞書Tが属する副分野の辞書とその主分野の辞書である。ヒットした複合語を含む複合語辞書Tが属する副分野の辞書とともにその主分野の辞書の優先順位も上位に上げることにより、副分野辞書のみではカバーできないその分野において一般的（全ての分野に一般的ではない）な語をカバーすることができ、分野を細分化したことにより生ずるデメリットを補うことができる。すなわち、分野の細分化によりその細分化された分野

（ここでは副分野）での訳質は向上する。しかし、分野が細分化されたゆえにその分野から外れると訳質は急激

に低下する恐れがある。しかし本実施の形態では、同時に大まかな分野の主分野辞書も上位に上げるので、細分化された副分野の周辺を補うことができる。よって、本実施の形態によりよりきめ細かな分野指定を可能にしつつ、ある程度の分野の幅を持たせて訳質の低下を抑制できる。なお、副分野辞書をさらに細分化して、その副分野の主分野辞書と副分野辞書を構成できる。

【0060】以上、本発明者によってなされた発明を発明の実施の形態に基づき具体的に説明したが、本発明は前記実施の形態に限定されるものではなく、その要旨を逸脱しない範囲で種々変更可能である。

【0061】たとえば、前記実施の形態では、辞書の自動切り換えのトリガになるデータとして複合語辞書を説明したが、図10に示すように、単語辞書を辞書の自動切り換えのトリガになる単語辞書Tと単語辞書を辞書の自動切り換えのトリガにならない単語辞書NTに分け、前記ステップ21において単語辞書Tに含まれる単語がヒットしたか否かを判断することができる。

【0062】また、前記実施の形態では辞書の自動切り換えのトリガになる複合語辞書Tについては優先順位が維持される例を説明した。これは、分野別辞書（あるいは主分野、副分野）の複合語辞書Tと基本辞書の複合語辞書とに同一の複合語が含まれる場合には基本辞書を選択するほうが妥当と考えたためであるが、このような重複がほとんど考えられない場合には、分野別辞書（あるいは主分野、副分野）の複合語辞書Tの優先順位を基本辞書よりも上位にしてもよい。

【0063】また、前記した未ヒット文数の設定値は10以外の値であってもよいのは勿論である。

【0064】また、前述のように、図4のフローは分野別辞書毎に実行されるので、図2のステップ11で入力された原文の翻訳中に2以上の分野別辞書の優先度が基本辞書より高くなることがある。例えば、スポーツ分野辞書の優先度が基本辞書より高い状態で、次の文に政治分野辞書の複合語辞書T（辞書切り換えのトリガになる複合語辞書）にヒットする複合語が現れたとき、政治分野別辞書の優先度も基本辞書より高くされる。その場合、後で複合語がヒットしたもののほど優先度を高くするのが好ましい。

【0065】

【発明の効果】本願で開示される発明のうち、代表的なものによって得られる効果は、以下の通りである。すなわち、より適切な辞書に自動切り換えが可能な翻訳システムを提供できる。より詳細な訳出に対応できる辞書構成を有する翻訳システムを提供できる。より訳質の高い翻訳文を生成できる。

【図面の簡単な説明】

【図1】本発明の一実施の形態を示す機械翻訳システムの概要を示すブロック図である。

【図2】図1の機械翻訳システムにおける翻訳処理の流

れの一例を示すフローチャートである。

【図 3】原文解析によって生成される構文木の一例を示す図である。

【図 4】辞書切り換え処理の一例を示すフローチャートである。

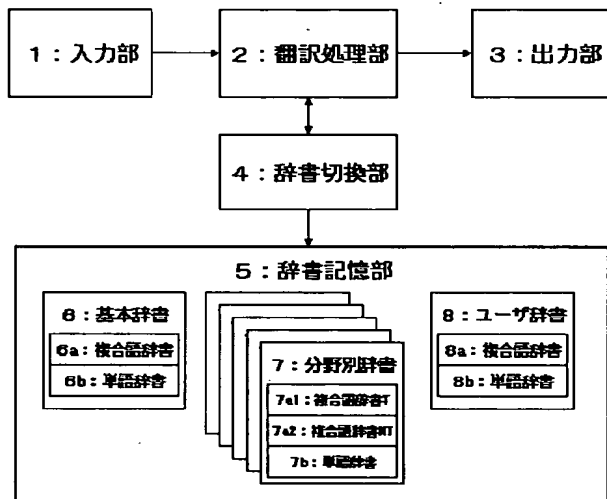
【図 5】実施の形態 1 の辞書構成の一例を示す図である。

【図 6】図 5 の辞書構成において優先順位が変更された後の辞書構成の一例を示す図である。

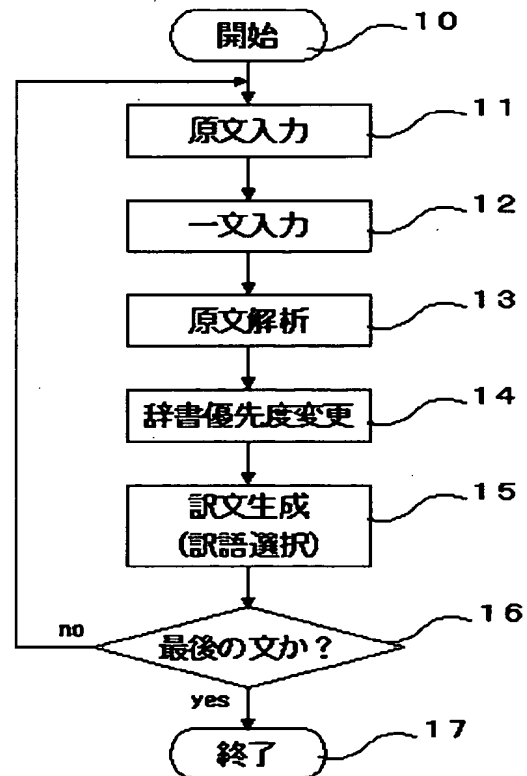
【図 7】実施の形態 2 の分野別辞書の構成の一例を示す 10
ブロック図である。

【図 8】実施の形態 2 の辞書構成の一例を示す図であ

【図 1】



【図 2】



る。

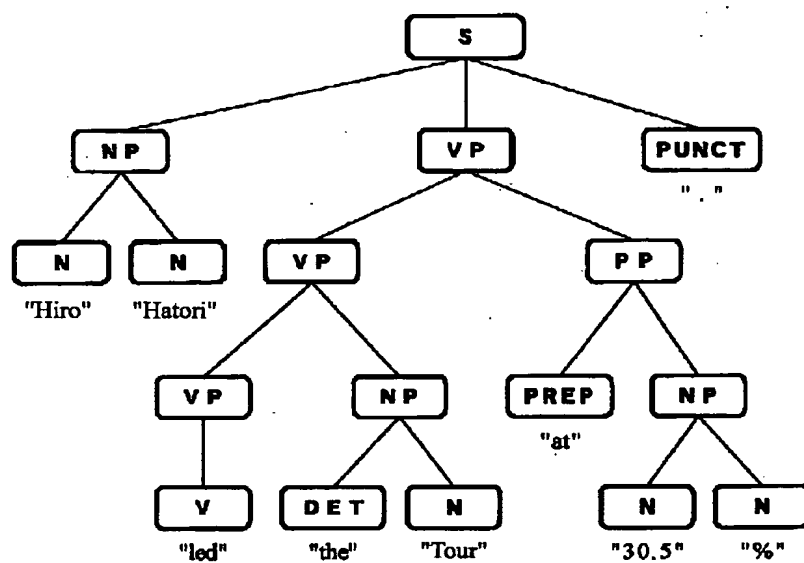
【図 9】図 8 の辞書構成において優先順位が変更された後の辞書構成の一例を示す図である。

【図 1 0】他の実施の形態の辞書構成を示す図である。

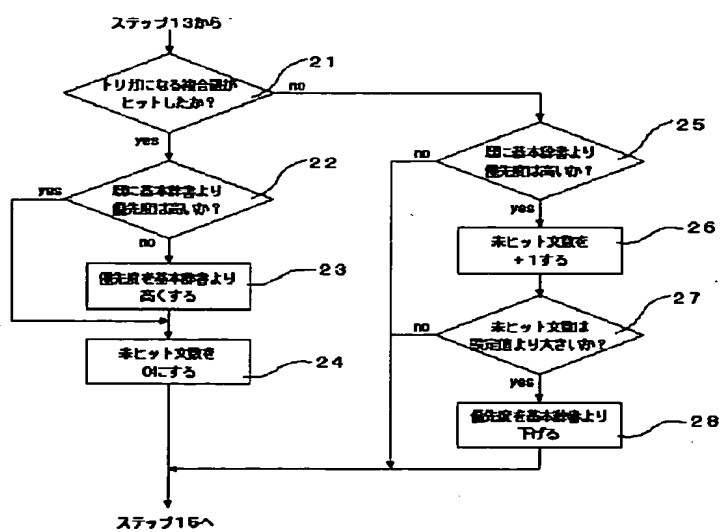
【符号の説明】

1…入力部、2…翻訳処理部、3…出力部、4…辞書切
換部、5…辞書記憶部、6…基本辞書、6 a…複合語辞
書、6 b…単語辞書、7…分野別辞書、7 a1…辞書切
換のトリガになる複合語辞書 (T)、7 a2…辞書切
換のトリガにならない複合語辞書 (NT)、7 b…単語辞書、8
…ユーザ辞書、3 6…主分野辞書、3 7…副分野辞書。

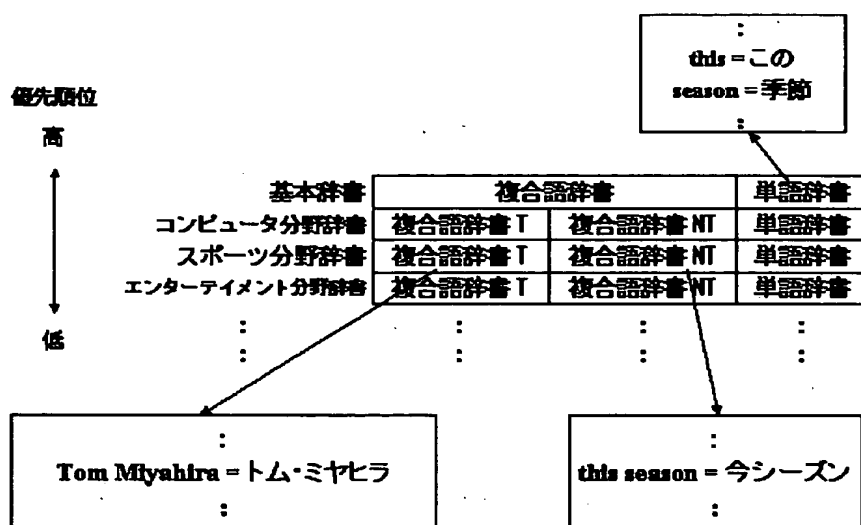
【図3】



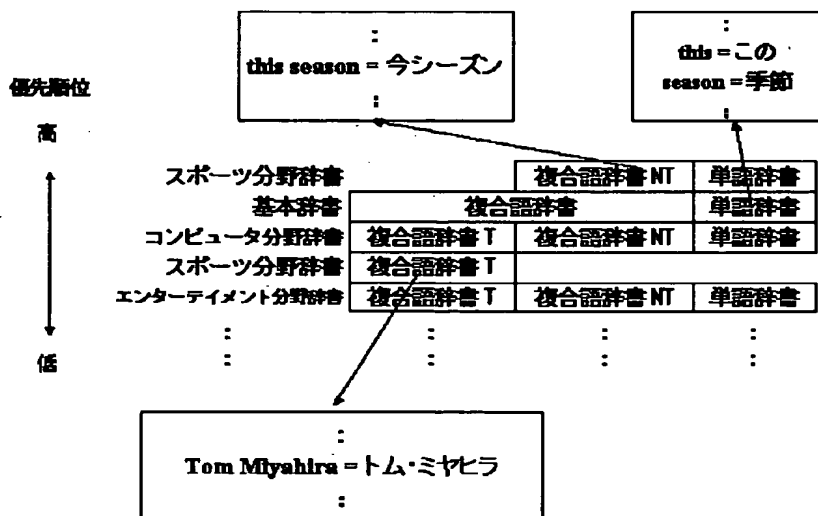
【図4】



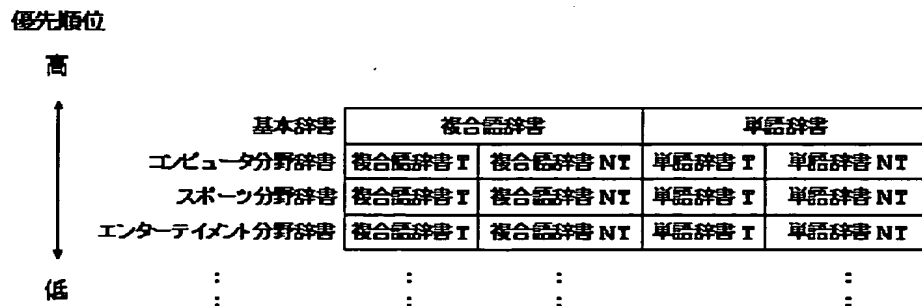
【図5】



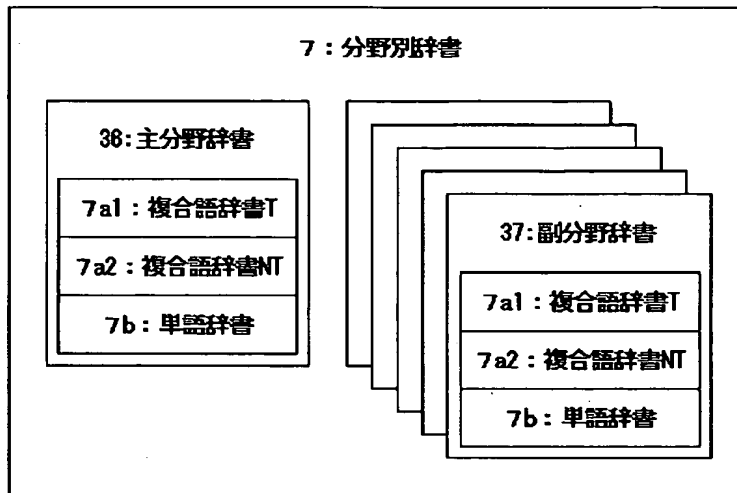
【図 6】



【图 10】



【図 7】

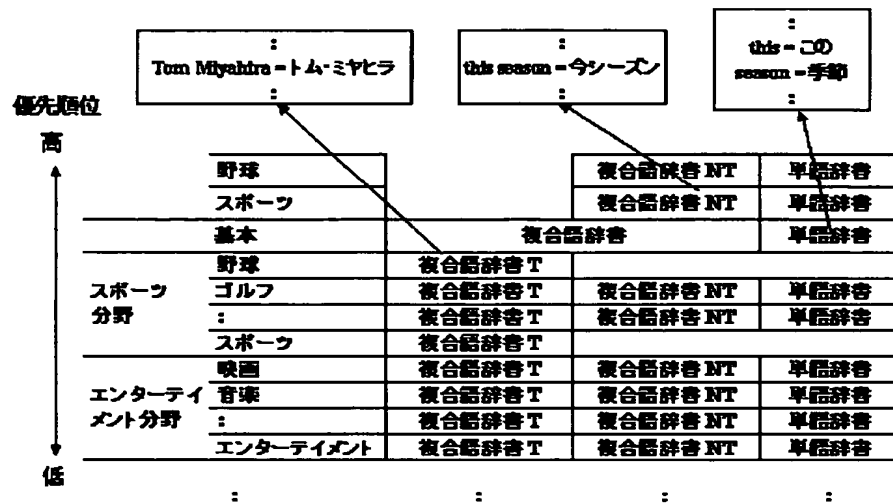


【図 8】

優先順位

	基本	複合語辞書		単語辞書
		複合語辞書 T	複合語辞書 NT	単語辞書
スポーツ分野	野球	複合語辞書 T	複合語辞書 NT	単語辞書
	ゴルフ	複合語辞書 T	複合語辞書 NT	単語辞書
	:	複合語辞書 T	複合語辞書 NT	単語辞書
	スポーツ	複合語辞書 T	複合語辞書 NT	単語辞書
エンターテインメント分野	映画	複合語辞書 T	複合語辞書 NT	単語辞書
	音楽	複合語辞書 T	複合語辞書 NT	単語辞書
	:	複合語辞書 T	複合語辞書 NT	単語辞書
	エンターテインメント	複合語辞書 T	複合語辞書 NT	単語辞書
高				
低	:	:	:	:

【図9】



フロントページの続き

(72)発明者 宮平 知博
神奈川県大和市下鶴間1623番地14 日本ア
イ・ビー・エム株式会社 大和事業所内

Fターム(参考) 5B091 AB06 AB11 CC02 CC14 CC15
EA07 EA17